

1 Legal action under Lebesgue integration

首先我们需要一些在 Lebesgue 积分下可以进行的操作，首先我们来看 **Fubini** 定理, 对于这个定理，我们需要明白下面两点，

- 它准许我们交换积分条件
- 该定理的使用条件是符号一致，或者 $f(X, Y)$ 是可积的

Theorem 1.1 X, Y is real valued, with induced measure μ_X, μ_Y , suppose $f: \mathcal{R}^2 \rightarrow \mathcal{R}$, *is non-negative or $\mathbb{E}[|f(X, Y)|] < \infty$* , then

$$\begin{aligned}\mathbb{E}[f(X, Y)] &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) d\mu_X(x) d\mu_Y(y) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) d\mu_Y(y) d\mu_X(x)\end{aligned}$$

通过 **Fubini** 定理, 我们可以证明如下论断: $Y \geq 0, \mathbb{E}[Y^p] = \int_0^{+\infty} py^{p-1} \mathbb{P}(Y > y) dy, p \geq 1$, 证明如下:

Proof For any $Z \geq 0$

$$z^p = \int_0^z py^{p-1} dy = \int_0^{+\infty} py^{p-1} 1_{\{y < z\}} dy$$

Now replace Z by $Y \geq 0$

$$\begin{aligned}\mathbb{E}[Y^p] &= \mathbb{E} \left[\int_0^{+\infty} py^{p-1} \mathbb{I}\{Y > y\} dy \right] \\ &\stackrel{\text{Fubini}}{=} \int_0^{+\infty} \mathbb{E}[p \cdot y^{p-1} \mathbb{I}\{Y > y\}] dy \\ &= \int_0^{+\infty} p \cdot y^{p-1} \mathbb{P}(Y > y) dy\end{aligned}$$

第二步之所以能够使用 **Fubini** 定理的是因为 $p \cdot y^{p-1} \mathbb{I}\{Y > y\}$ 始终大于 0。
请注意，如果使用分部积分的方法，你需要考虑一个 $\lim_{x \rightarrow \infty} y^p \mathbb{P}(Y > y)$ 的式子，这个式子在 $\mathbb{E}(Y^p) = \infty$ 时不一定趋于 0，例如可以验证 Cauchy 分布的一阶中心距。所以如果要用分部积分的方法，还需要分 $\mathbb{E}[Y^p]$ 是否趋于无穷，这是很麻烦的 (*况且我还不知道如果 $\mathbb{E}[Y^p]$ 趋于 ∞ 怎么做*)

不过在这里值得一提的是, 如果 $\mathbb{E}(Y^p) < \infty$, 我们可以知道 $\lim_{x \rightarrow \infty} y^p \mathbb{P}(Y > y) \rightarrow 0$, 证明如下

Proof For any $y \geq 0$

$$y^P \mathbb{I}\{Y > y\} \leq \mathbb{E} Y^P \mathbb{I}\{Y > y\} \leq Y^P$$

because

$$Y^P \mathbb{I}\{Y > y\} \xrightarrow{P} 0$$

then by DCT

$$\mathbb{E} [y^P \mathbb{I}\{Y > y\}] \leq \mathbb{E} [Y^P \mathbb{I}\{Y > y\}] = 0$$

在这里第二个式子和第三个式子用到了 p 阶矩小于 ∞ 的条件。

再再再多说一点, 即使 $\mathbb{E}(Y^p) = \infty$, 我们也不能判断 $\lim_{x \rightarrow \infty} y^p \mathbb{P}(Y > y)$ 是否趋于无穷, 这个式子也是可以趋于 0 的, 比如可以考虑如下的概率分布函数, $P(X_i > x) = e/x \log x$ 。

总结一下这里讨论的内容

- 无论在什么情况下 $Y \geq 0, \mathbb{E}[Y^p] = \int_0^{+\infty} p y^{p-1} \mathbb{P}(Y > y) dy, p \geq 1$
- $\mathbb{E}(Y^p) < \infty$, 则 $\lim_{x \rightarrow \infty} y^p \mathbb{P}(Y > y) \rightarrow 0$
- $\lim_{x \rightarrow \infty} y^p \mathbb{P}(Y > y) \rightarrow 0$, 也不能判断 $\mathbb{E}(Y^p)$ 是否小于无穷

未完待续... 需要补充极限和积分交换的定理 (控制收敛定理等)

2 大数定律的证明和运用

在这里我们介绍大数定律的基本证明和运用方式, 不同于高等统计, 在这里我们关心 why (in pure math) and how (in example with more mathematical meaning)

2.1 L^2 Weak Laws

这里的证明思路实在是太 normal 了, 只要计算方差即可, 不过这也是证明依概率收敛的一贯思路。在这里不再多说了

2.2 弱大数律 (SLLN)

这里讨论了一种 triangular 的随机变量列 (**Tiangular Arrays**), 我觉得就是随着 n 的增加, 对 $S_n = X_{n,1} + \cdots + X_{n,n}$ 中的元素加一个随着 n 变化的条件罢了, 可以对比一下之前大多数都是 X_i 的需要的性质都是固定的

$$\begin{array}{cccc} X_{1,1} & & & \\ X_{2,1} & X_{2,2} & & \\ X_{3,1} & X_{3,2} & X_{3,3} & \\ \vdots & & & \\ X_{n,1} & X_{n,2} & \cdots & X_{n,3} \end{array}$$

接下来依据这个概念，我们来介绍相关定理

Theorem 2.1 *Let $\mu_n = ES_n, \sigma_n^2 = \text{var}(S_n)$. If $\sigma_n^2/b_n^2 \rightarrow 0$ then*

$$\frac{S_n - \mu_n}{b_n} \rightarrow 0 \quad \text{in probability}$$

Proof 这个的证明也比较 *normal*, 只要计算方差就行

Remark 2.1.1 运用这个定理主要的困难就是计算方差来找到 b_n , 想要得到一个更加实用的结果的话我们需要找到一个 b_n , 满足 $\frac{\mu_n}{b_n} \rightarrow c$, 这样我们可以得到 $\frac{S_n}{b_n} \xrightarrow{P} c$, 具体的例子我们将在后面讨论。

接下来我们考虑一个比较经典的方法，**截断法 (Truncation)**，这意味着考虑这样的随机变量

$$\bar{X} = X1_{(|X| \leq M)} = \begin{cases} X & \text{if } |X| \leq M \\ 0 & \text{if } |X| > M \end{cases}$$

接下里我们可以考虑截断法下的 triangular r.v.

Theorem 2.2 (Weak law for triangular arrays.) *For each n let $X_{n,k}, 1 \leq k \leq n$, be independent. Let $b_n > 0$ with $b_n \rightarrow \infty$, and let $\bar{X}_{n,k} = X_{n,k}1_{(|X_{n,k}| \leq b_n)}$. Suppose that as $n \rightarrow \infty$*

- (i) $\sum_{k=1}^n P(|X_{n,k}| > b_n) \rightarrow 0$, and
(ii) $b_n^{-2} \sum_{k=1}^n E \bar{X}_{n,k}^2 \rightarrow 0$.

If we let $S_n = X_{n,1} + \dots + X_{n,n}$ and put $a_n = \sum_{k=1}^n E\bar{X}_{n,k}$ then $(S_n - a_n)/b_n \rightarrow 0$ in probability

Proof 证明也是十分简单的, 令 $\bar{S}_n = \bar{X}_{n,1} + \cdots + \bar{X}_{n,n}$, 只需要注意到分解

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right) \leq P(S_n \neq \bar{S}_n) + P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right)$$

前一项因为条件 (i) 规定的尾概率趋于 0, 后一项被条件 (ii) 规定的方差控制住趋于 0.

Remark 2.2.1 这个定理的说明了只要尾概率不是太大, 就能被控制住, 当然这里解决了定理 2.1 中 μ_n 不存在的问题, 你可以把 μ_n 替换成 *truncation* 之后的 a_n 就行了, 至于一个合适的 b_n 所需要的和定理 2.1 的 *remark* 中讨论的一样

有了这个定理之后, 我们可以得到关于 i.i.d 的弱大数定理的一个比较优美的条件, 这个条件也比较符合直观, 即只要尾概率不是太大, 就可以收敛

Theorem 2.3 (Weak law of large numbers) *Let X_1, X_2, \dots be i.i.d. with*

$$xP(|X_i| > x) \rightarrow 0 \quad \text{as } x \rightarrow \infty$$

Let $S_n = X_1 + \cdots + X_n$ and let $\mu_n = E(X_1 1_{(|X_1| \leq n)})$. Then $S_n/n - \mu_n \rightarrow 0$ in probability.

Proof 我们只需要验证定理 2.2 中的两个条件就可以了, 这里很显然 $b_n = n$ 对于 (i) $\Leftrightarrow nP(|X_1| > n)$, 这是显然成立的。

对于 (ii), 利用之前的论断 $Y \geq 0, \mathbb{E}[Y^p] = \int_0^{+\infty} py^{p-1}\mathbb{P}(Y > y)dy, p \geq 1$, 我们有

$$E(\bar{X}_{n,1}^2)/n = \frac{1}{n} \int_0^n 2yP(|X_1| > y) dy$$

因为这相当于对 $yP(|X_1| > y)$ 求平均, 而这个趋于 0, 所以我们可以验证 (ii)。具体的证明手段可以分成两种

1. 对于 $\forall \epsilon$, 分成两段, 当 $x > M$ 时, $xP(|X_i| > x) < \epsilon$, 当 $x < M$ 时, 利用 $\frac{1}{n}$ 去控制
2. 或者直接 $g(y) = 2yP(|X_1| > y)$ 且 $g_n(y) = g(ny)$, 则 $\lim_{n \rightarrow \infty} g_n(y) = 0$, 之后利用控制收敛定理

$$(1/n) \int_0^n g(y)dy = \int_0^1 g_n(x)dx \rightarrow 0$$

Remark 2.3.1 定理2.3中的假设是使得满足 $\frac{S_n}{b_n} - a_n \rightarrow 0$ 存在的必要条件, 具体可见 *Feller(1971)*

接下来我们来举一个例子来具体说明 **remark2.1.1**中的意思以及大数定律一些更广泛的运用

Example 2.1 假设 X_1, X_2, \dots 是一些独立的随机变量列满足

$$P(X_i = 2^j) = 2^{-j}$$

可以想象现在有一场赌博游戏, 你可以得到 2^j 元如果你扔出 j 次硬币猜得到正面, 很容易观察到 $EX_1 = \infty$, 这样我们没法直接要求赌徒每局都支付这个随机变量的期望的钱数, 但是利用定理2.2, 我们可以得到 S_n 的渐近行为, 并得到一个与 n 有关的需要付的钱数

即我们现在需要构造

$$\frac{S_n - a_n}{b_n} \rightarrow 0$$

当然我们希望在挑选一个合适的 b_n , 使得 $\frac{a_n}{b_n}$ 趋于一个常数 c , 这样我们可以得到 $\frac{S_n}{cb_n} \rightarrow 1$, 这样我们按照 n 局游戏收费 cb_n 元也算公平。在这里我们需要注意两点

- 一个过于大的 b_n 是没有用的, 例如大到 $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$ 致使我们最后得到 $\frac{S_n}{b_n} \rightarrow 0$ 这是没啥信息的, 我们只知道 $S_n = o(b_n)$
- 这里的“公平”和期望那种“公平”还是有些许不一样, 毕竟期望那种是相减是 0 , $\frac{S_n}{n} - \mu_n \rightarrow 0$, 最后即使亏一块钱的概率都是很小的 (?), 但是这里是比值趋于 1 , 这个结论更多的是在描述速率, 这个结论推不出上面的相减为 0 , (反例是显然的), 但是相减为 0 加上 $\lim_{n \rightarrow \infty} b_n \neq 0$ 可以推出相除为 1

所以我们为了找到这样的 b_n , 我们可以去代入定理2.2中的条件 (i)(ii) 再加上 $\frac{a_n}{b_n} \rightarrow 1$, 为了方便, 我们可以记 $b_n = 2^m$, 则满足 (i)(ii) $\Leftrightarrow 2^{-m}n \rightarrow 0$, 计算得到 $a_n = nm$, 所以 $\frac{a_n}{b_n} \rightarrow 1 \Leftrightarrow \frac{nm}{2^m} \rightarrow 1$, 总的来说我们需要找到这样一个 m , 使得

$$\frac{n}{2^m} \rightarrow 0 \tag{1}$$

$$\frac{nm}{2^m} \rightarrow 1 \tag{2}$$

令 $n \cdot 2^{-m} = s(n)$, 则 $m = \log n - \log s(n)$, 那么条件 (2) 就变成了

$$\frac{s(n)[n \cdot \log(n) - \log s(n)]}{n} \rightarrow 1 \quad s(n) \rightarrow 0$$

观察一下, 取 $s(n) = \frac{1}{\log n}$ 即可, 不过这里可以再进一步精细一下使得 m 是个整数, 强行用 \sup 规定一下就行, 再代入原式我们就可以得到

$$\frac{\sum_{k=1}^n X_k}{n \log n} \rightarrow 1$$

这里都是以 2 为底的 \log , 当一个赌徒玩 n 局游戏时, 我们可以收 $n \log n$ 的费用来保证”公平”

2.3 强大数律 (SLLN)

首先我们一上来就介绍的是 **Borel-Cantelli lemmas**, 在介绍这个 lemma 之前, 我们需要辨析几个极限的概念

首先是事件的极限, 分为事件的上极限和下极限,

Def 2.1 事件的上极限的定义是

$$\limsup A_n = \bigcap_{m \rightarrow \infty} \bigcup_{n=m}^{\infty} A_n = \lim_{m \rightarrow \infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}$$

事件的下极限的定义是

$$\liminf A_n = \bigcup_{m \rightarrow \infty} \bigcap_{n=m}^{\infty} A_n = \lim_{m \rightarrow \infty} \bigcap_{n=m}^{\infty} A_n = \{\omega \text{ that are in all but finitely many } A_n\}$$

翻译一下这两句英文, 上极限的样本点就是指那些会无限出现在事件列的样本点, 但可能也有无穷多的事件不包含这些样本点哦; 下极限的样本点就是指那些在某个事件之后 (maybe 很大的 M), 所有的事件都包含的样本点。

接下来我们对比一下数列的极限和随机变量的极限, 避免之后犯迷糊。

Def 2.2 数列 $\{a_n\}$ 的上极限是指

$$\limsup x_n = \inf_{n \geq 0} \sup_{k \geq n} x_k = \inf \{\sup \{x_k : k \geq n\} : n \geq 0\}$$

或者

$$\limsup x_n = \lim_{n \rightarrow \infty} \left(\sup_{m \geq n} x_m \right)$$

数列 $\{a_n\}$ 的下极限是指

$$\liminf x_n = \sup_{n \geq 0} \inf_{k \geq n} x_k = \sup \{\inf \{x_k : k \geq n\} : n \geq 0\}$$

或者

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \left(\inf_{m \geq n} x_m \right)$$

至于随机变量的上下极限，就是你把 $X_n(\omega)$ 看成数列，然后照数列的上下极限处理就行。可以看到这些上下极限的定义是很像的，但是还是要看清到底是随机变量的极限还是事件的极限，不然有些地方理解起来特别难受。

接下来我们需要介绍 Borel-cantelli 引理，

Theorem 2.4 (Borel-Cantelli lemma) *If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then*

$$P(A_n \text{ i.o.}) = 0.$$

] If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then

$$P(A_n \text{ i.o.}) = 0.$$

Proof 证明只需要计算 $N = \sum_k 1_{A_k}$ 的期望即可。

或者我们给出一个 *Intuitive* 的想法，如果 $P(A_n \text{ i.o.}) \neq 0$ ，那么就有一个有概率测度的事件/区间 B_n 一直出现在事件列中，那么 $P(A_n) \geq \infty \times P(B_n) = \infty$ ，这肯定不成立了，**所以有的时候我们只需要去说明 $P(A_n)$ 收敛的够快就行了**

Remark 2.4.1 这是一个涉及到了事件上极限的引理，如果咱们回顾几乎处处收敛的定义，我们会发现他就是用数列上极限定义的，所以这个定理主要是用来证明几乎处处收敛的，**(截止到第 6 次课，Borel-Cantelli 引理和单调的随机变量列的“两面夹芝士”法是唯二的证明几乎处处收敛的方法)**

下面这个例子是一个利用 Borel-Cantelli 引理证明几乎处处收敛的例子，并且用到了上面说的收敛够快的思想

Theorem 2.5 $X_n \rightarrow X$ in probability if and only if for every subsequence $X_{n(m)}$ there is a further subsequence $X_{n(m_k)}$ that converges almost surely to X .

Proof Let ϵ_k be a sequence of positive numbers that $\downarrow 0$. For each k , there is an $n(m_k) > n(m_{k-1})$ so that $P(|X_{n(m_k)} - X| > \epsilon_k) \leq 2^{-k}$. Since

$$\sum_{k=1}^{\infty} P(|X_{n(m_k)} - X| > \epsilon_k) < \infty$$

the Borel-Cantelli lemma implies $P(|X_{n(m_k)} - X| > \epsilon_k \text{ i.o.}) = 0$, i.e., $X_{n(m_k)} \rightarrow X$ a.s. To prove the second conclusion, we note that if for every subsequence $X_{n(m)}$ there is a further subsequence $X_{n(m_k)}$ that converges almost surely to X then we can apply the next lemma to the sequence of numbers $y_n = P(|X_n - X| > \delta)$ for any $\delta > 0$ to get the desired result.

这里相当于给 $P(|X_{n(m_k)} - X|)$ 设定的收敛速度是 2^{-k} ，当然这里构造了一个向下递减的 ϵ_k 也很有意思，如果你根据一个固定的 ϵ 去挑选 m_k ，那么每个 ϵ 挑选出来的 m_k 是不一样的，这样就无法说 ϵ 的任意性了

Theorem 2.6 Let X_1, X_2, \dots be i.i.d. with $EX_i = \mu$ and $EX_i^4 < \infty$. If $S_n = X_1 + \dots + X_n$ then $S_n/n \rightarrow \mu$ a.s.

Proof 这里就是利用四阶矩给 $P(|S_n| > n\epsilon)$ 构造了一个 $\frac{1}{n^2}$ 的收敛速率

在这个定理阐述的最后，我再多提一件我在课上讨论出来但是回顾的时候发现也没有用到的讨论，那就是证明 $P(A_n \text{ i.o.}) = 0$ 能证明这个事件在 n 足够大时几乎不发生，但是 $P(A_n \text{ i.o.}) = 1$ ，却不能说明 n 足够大时，后面的都发生。

接下来我们介绍 Borel-Cantelli 第二引理，这个引理对 $P(A_n \text{ i.o.}) = 1$ 给出了一个充分条件

Theorem 2.7 (The second Borel-Cantelli lemma) If the events A_n are independent then $\sum P(A_n) = \infty$ implies $P(A_n \text{ i.o.}) = 1$.

Proof 主要是利用 $1-x \leq e^{-x}$ ，弄成 $\prod_{n=M}^N (1 - P(A_n)) \leq \prod_{n=M}^N \exp(-P(A_n))$ ，这样就转换到了 $\sum P(A_n)$ 上

Remark 2.7.1 这个定理的基本思想是这样的，如果有一个事件/区间 B_n 不属于 A_n 的上极限，这就意味着在某个 m 之后一直到无穷，他就再也不出现了，他的概率起码小于 $P(\cap_{n=M}^{\infty} A_n^c)$ ，因为规定了 $\sum P(A_n) = \infty$ 的这样一个速率，所以直观上前面这个概率 $P(\cap_{n=M}^{\infty} A_n^c)$ 越来越小趋于 0

下面这个定理是 Borel-Cantelli 引理的应用，他揭示了大数定律中 $E|X_i| \leq \infty$ 对于 μ_n 存在是必要的。

Theorem 2.8 If X_1, X_2, \dots are i.i.d. with $E|X_i| = \infty$, then $P(|X_n| \geq n \text{ i.o.}) = 1$. So if $S_n = X_1 + \dots + X_n$ then $P(\lim S_n/n \text{ exists} \in (-\infty, \infty)) = 0$.

Proof 利用

$$E|X_1| = \int_0^\infty P(|X_1| > x) dx \leq \sum_{n=0}^\infty P(|X_1| > n)$$

加上 Borel-Cantelli 第二引理得到 $P(|X_n| \geq n \text{ i.o.}) = 1$, 接着导出

$$\left| \frac{S_n}{n} - \frac{S_{n+1}}{n+1} \right| > 2/3 \quad \text{i.o.}$$

那么这个集合和 $C \equiv \{\omega : \lim_{n \rightarrow \infty} S_n/n \text{ exists} \in (-\infty, \infty)\}$ 的交集就是空集, 这就导出了 $P(C) = 0$

下面这个定理不仅说明了 $P(A_n \text{ i.o.}) = 1 (\Leftrightarrow \sum_{m=1}^\infty 1_{A_m} = \infty)$, 还说明趋于 ∞ 的速度

Theorem 2.9 (两面夹芝士法) *If A_1, A_2, \dots are pairwise independent and $\sum_{n=1}^\infty P(A_n) = \infty$ then as $n \rightarrow \infty$*

$$\sum_{m=1}^n 1_{A_m} / \sum_{m=1}^n P(A_m) \rightarrow 1 \quad \text{a.s.}$$

Proof 我们分三步梳理一下这个证明

Step 1 说明 $\sum_{m=1}^\infty 1_{A_m} = \infty$ 依概率收敛, 这一步是为了给后面安插概率方便

step 2 令 $n_k = \inf\{n : ES_n \geq k^2\}$ and $T_k = S_{n_k}$, 因为这是示性函数, 所以我们有 $k^2 \leq ET_k \leq k^2 + 1$, 那么我们可以得到

$$P(|T_k - ET_k| > \delta ET_k) \leq 1/(\delta^2 k^2)$$

这样我们可以得到 T_k 几乎处处收敛

Step 3 接下来利用递增这个性质, 我们可以得到 $\frac{T_k(\omega)}{ET_{k+1}} \leq \frac{S_n(\omega)}{ES_n} \leq \frac{T_{k+1}(\omega)}{ET_k}$ 这样就把 $\frac{S_n(\omega)}{ES_n}$ “夹”住了, 接下来我们做一步转换

$$\frac{ET_k}{ET_{k+1}} \cdot \frac{T_k(\omega)}{ET_k} \leq \frac{S_n(\omega)}{ES_n} \leq \frac{T_{k+1}(\omega)}{ET_{k+1}} \cdot \frac{ET_{k+1}}{ET_k}$$

接下来我们只需要证明 $ET_{k+1}/ET_k \rightarrow 1$ 即可, 这是显然的因为

$$k^2 \leq ET_k \leq ET_{k+1} \leq (k+1)^2 + 1$$

Remark 2.9.1 对于这个定理，我们从意义和证明方法上给一些注释

1. 根据大数定律, $\sum_{m=1}^n 1_{A_m} / \sum_{m=1}^n P(A_m) (1 - P(A_m)) \rightarrow \infty$
2. $\sum_{m=1}^n 1_{A_m}(\omega)$ 这个东西可以看做是样本点 ω 在事件列中出现的频率, 这个定理相当于给这个频率进行了一个刻画
3. 相比于定理 2.5, 这里挑选 n_k 的方式更加细腻, 有点像手动挑选的感觉, 这是为了给 ET_k 找到上下界, 之后才能证明 $ET_{k+1}/ET_k \rightarrow 1$
4. 对于一个递增的数列 X_n , 我们想要证明 $X_n/c_n \rightarrow 1$, 我们需要找到 c_{nk} where $c_{n(k+1)}/c_{n(k)} \rightarrow 1$
5. 这里能找到这样的 $c_{n(k)}$, 有一个上界 $(k+1)^2 + 1$, 主要运用了示性函数和 PA_m 都不大于 1 的特性, 像在 **Exercise 2.3.19** 中, 就需要泊松分布的可加性来使得 $\lambda_i < 1$ 来保证步子不会迈的太大

接下来我们正式证明强大数律

Theorem 2.10 (Strong law of large numbers) *Let X_1, X_2, \dots be pair-wise independent identically distributed random variables with $E|X_i| < \infty$. Let $EX_i = \mu$ and $S_n = X_1 + \dots + X_n$. Then $S_n/n \rightarrow \mu$ a.s. as $n \rightarrow \infty$.*

Proof 主要的证明步骤也是三步 (怎么都是三步, 很难不让人怀疑这不是在嗯蹭欧几里得的三段式, $Y_k = X_k \mathbb{I}\{|X_k| < k\}$)

Step 1 我们需要证明 $\frac{|S_n - T_n|}{n} \rightarrow 0$ a.s., 这是一个 $P(X_k \neq Y_k \text{ i.o.}) = 0$ 的一个推论

step 2 接下来我们需要证明这么一个东西

$$\sum_{k=1}^{\infty} \text{var}(Y_k)/k^2 \leq 4E|X_1| < \infty$$

这个东西看上去还是很神奇的, 毕竟是低阶矩控制住了高阶矩, 但实际上我们可以用离散分布中的一个点来给一个比较 *intuitive* 的观点, 假设在一个离散分布中 $P(X = k) = p_k$, 那么在 $\sum_{k=1}^{\infty} \text{var}(Y_k)$ 这个 k 的贡献就是

$$\sum_{K > k}^{\infty} p_k k^2 \cdot \frac{1}{K^2} = p_k k \sum_{K > k}^{\infty} \frac{k}{K^2}$$

那么后面这个 *summation* 求和出来肯定是被 *bound* 住的，你再把所有的 k 都求和就能更清晰了。

再往深点说这就是那个技巧：当 $X < M$ 时， $E(X^2) < ME(X)$

再再再往深点说，你把这个 X 压到 $[0, 1]$ 之间，再平方是会变小的，不过小的数求和多，大的数求和少就就会被一阶矩控制住（其实你把连续按照整数点分段就和离散的情况一样了）

Step 3 接下来就是经典的两面夹芝士法了，关键就是找到这个子列是吧，这里找的这个子列 $k(n) = \lceil \alpha^n \rceil$ ，找这个子列有什么好处呢，就是你可以用上下极限去控制 $\frac{k(n)}{k(n+1)}$ ，而不是直接 $\frac{k(n)}{k(n+1)} = 1$ ，好吧这个东西还是很妙的，不仅控制住了 *summation*，还控制住了 $\frac{k(n)}{k(n+1)}$

当然这里我们可以取别的 $k(n)$ 也可以做，（我个人觉得），例如取 $k(n) = \lceil \frac{n^2}{EX} \rceil + 1$ ，这样我们可以 *bound* 住 $k(n)$ by $\frac{n^2}{EX} - 1 \leq k(n) \leq k(n+1) \leq \frac{(n+1)^2}{EX} + 1$ ，当然这实际上是取了

$$k(n) = \inf\{k : E(S_k) > n^2\}$$

!!!! 注意这里标红的地方全写错了，根本原因是计算错了 $\sum_{n:k(n)>j} \frac{1}{k(n)^2}$ 如果这样取的话这个 *summation* 算出来 $\sim \frac{1}{j}$ ，你把 $k(n)$ 整体考虑，你就会发现这个取多项式肯定是不行的

总的来说，这个 $k(n)$ 的选取应该满足两个条件

- 可以控制住 $\frac{k(n+1)}{k(n)}$
- $\sum_{n:k(n)>j} \frac{1}{k(n)^2}$ 至少要是 j^{-2} 的阶数

所以这个取法还是很妙的，暂时也没有什么 *intuitive* 的解释，之后无非就是利用上下极限来控制住 $\frac{k(n+1)}{K(n)}$

接下来，依托大数定律讲了三个例子和应用，也是比较常见的，首先是更新定理的一个例子

Example 2.2 (Renewal theory) Let X_1, X_2, \dots be i.i.d. with $0 < X_i < \infty$. Let $T_n = X_1 + \dots + X_n$ and think of T_n as the time of n th occurrence of some event. For a concrete situation, consider a diligent janitor who replaces a light bulb the instant it burns out. Suppose the first bulb is put in at time 0 and let X_i be the lifetime of the i th light bulb. In this interpretation,

T_n is the time the n th light bulb burns out and $N_t = \sup \{n : T_n \leq t\}$ is the number of light bulbs that have burnt out by time t .

If $EX_1 = \mu \leq \infty$ then as $t \rightarrow \infty$,

$$N_t/t \rightarrow 1/\mu \quad a.s. \quad (1/\infty = 0)$$

这个东西只需要观察到 $\frac{T(N_t)}{N_t} \leq \frac{t}{N_t} \leq \frac{T(N_t+1)}{N_t+1} \cdot \frac{N_t+1}{N_t}$, 然后两边用一下大数定律一夹就 OK 了

下一个定理说明的是经验过程的收敛性

Theorem 2.11 (The Glivenko-Cantelli theorem) Let X_1, X_2, \dots be i.i.d. with distribution F and let

$$F_n(x) = n^{-1} \sum_{m=1}^n 1_{(X_m \leq x)}$$

Then as $n \rightarrow \infty$, $\sup |F_n(x) - F(x)| \rightarrow 0 \quad a.s.$

Proof 证明这个, 首先要做的是分点, 记 $x_{j,k} = \inf\{y : F(y) \geq j/k\}$, 那么当 n 足够大时

$$|F_n(x_{j,k}) - F(x_{j,k})| < k^{-1} \quad \text{and} \quad |F_n(x_{j,k-}) - F(x_{j,k-})| < k^{-1}$$

在这里之所以要引入右极限是因为, 只用第一个界是控制不住的, 因为我们没发控制 $|F_n(x_{j,k}) - F_n(x_{j-1,k})|$ 导致没法控制住 $F(x)$, 引入右极限之后我们可以得到

$$F(x_{j,k-}) - F(x_{j-1,k}) \leq k^{-1}$$

这样的话我们就可以通过 $F(x_{j-1,k}) \leq F(x) \leq F(x_{j,k-})$ 来控制了

2.4 随机序列的收敛性

其实我也不懂什么叫随机序列的收敛, 感觉这一节就是介绍了一些关于科尔戈莫洛夫 (Kolmogorov) 的一些结论, 然后从另一方面证明了强大数律。首先是这个神奇的定理

Theorem 2.12 (Kolmogorov's maximal inequality) Suppose X_1, \dots, X_n are independent with $EX_i = 0$ and $\text{var}(X_i) < \infty$. If $S_n = X_1 + \dots + X_n$ then

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right) \leq x^{-2} \text{var}(S_n)$$

这个定理神奇就神奇在右边这个界通常是 $P(|S_n| > x)$ 的，但是我们发现右边这个界还可以对所有的 $|S_k|, k < n$ 都进行控制。

Proof 证明相当于是计算 ES_n^2 的时候把它分到 n 个不交的事件 $A_k = \{|S_k| \geq x \text{ but } |S_j| < x \text{ for } j < k\}$ 上计算，然后在这 n 个不交的时间上我们做分解 $(S_n - S_k + S_k)^2$ ，我们发现在这些时间上交叉项算出来是 0（有点像鞅），主项算出来是

$$\sum_{k=1}^n \int_{A_k} S_k^2 dP \geq \sum_{k=1}^n x^2 P(A_k) = x^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq x\right)$$

然后把 x^2 除过去就行了

接下来我们要得到一个关于序列收敛的定理，当然这里可以是收敛到无穷，证明的手法是柯西收敛定理

Theorem 2.13 Suppose X_1, X_2, \dots are independent and have $EX_n = 0$. If

$$\sum_{n=1}^{\infty} \text{var}(X_n) < \infty$$

then with probability one $\sum_{n=1}^{\infty} X_n(\omega)$ converges.

Proof 这个东西的证明思路就是 $w_M = \sup_{m, n \geq M} |S_m - S_n|$ 趋于 0 almost surely，再详细一点，（简易理解版，具体还是极限）这就是说存在一个空间 $\Omega_0, P(\Omega_0) = 1$ ，在这个空间上所有的样本点上 $w_M \rightarrow 0$ ，这就代表着这个空间样本点上，这个 summation 都是收敛的。证明 $w_M = \sup_{m, n \geq M} |S_m - S_n|$ 趋于 0 almost surely 的思路就是，首先利用定理 2.12 得到

$$P\left(\sup_{m \geq M} |S_m - S_M| > \epsilon\right) \leq \epsilon^{-2} \sum_{n=M+1}^{\infty} \text{var}(X_n) \rightarrow 0 \quad \text{as } M \rightarrow \infty$$

然后利用 $w_M < 2 \sup_{m \geq M} |S_m - S_M|$ ，我们就可以得到

$$P(w_M > 2\epsilon) \leq P\left(\sup_{m \geq M} |S_m - S_M| > \epsilon\right) \rightarrow 0$$

然后我们利用 w_M 递减的性质，我们就可以得到

$$P(\cap_{m \rightarrow \infty} \cup_{n=m} \{w_n > \epsilon\}) = \lim_{m \rightarrow \infty} P(\cup_{n=m} \{w_n > \epsilon\}) = \lim_{m \rightarrow \infty} P\{w_m > \epsilon\} = 0$$

接下来是一个很像强大数律的定理，但是得到的结论是收敛，不过和下面那个 **Kronecker's Lemma** 结合，我们可以很快得到一个弱化版的强大数律

Theorem 2.14 (Kolmogorov's three-series theorem) *Let X_1, X_2, \dots be independent. Let $A > 0$ and let $Y_i = X_i 1_{(|X_i| \leq A)}$. In order that $\sum_{n=1}^{\infty} X_n$ converges a.s., it is necessary and sufficient that (i) $\sum_{n=1}^{\infty} P(|X_n| > A) < \infty$ (ii) $\sum_{n=1}^{\infty} EY_n$ converges, $\sum_{n=1}^{\infty} \text{var}(Y_n) < \infty$*

这个定理有了前面的铺垫之后，充分性就是显然的，必要性之后再证。接下来是这个 **Kronecker's Lemma**

Theorem 2.15 (Kronecker's lemma) *If $a_n \uparrow \infty$ and $\sum_{n=1}^{\infty} x_n/a_n$ converges then*

$$a_n^{-1} \sum_{m=1}^n x_m \rightarrow 0$$

Proof 如果我们令 $b_m = \sum_{k=1}^m \frac{x_k}{a_k}$ ，我们就可以得到

$$a_n^{-1} \sum_{m=1}^n x_m = b_n - \sum_{m=1}^n \frac{a_m - a_{m-1}}{a_n} b_{m-1}$$

把后面这个东西看成一个加权求和，下标较大的部分 b_{m-1} 接近于极限，权重求和为 $1 - \frac{a_M}{a_n}$ 接近于 1（因为 $a_n \uparrow \infty$ ），下标较小的部分权重求和接近于 0，这就证明完了

这个定理的意思就是你把收敛的数列的分母都换成最大的那个就会趋于 0，然后利用这个定理，我们就可以得到下面这个需要 iid 的大数定律

Theorem 2.16 (The strong law of large numbers) *Let X_1, X_2, \dots be i.i.d. random variables with $E|X_i| < \infty$. Let $EX_i = \mu$ and $S_n = X_1 + \dots + X_n$. Then $S_n/n \rightarrow \mu$ a.s. as $n \rightarrow \infty$.*

Proof 证明前面 *truncation* 的东西都是和原始那个一样的，那如果我们令 $Z_k = \frac{Y_k - EY_k}{k}$ 的话，我们可以计算方差得到 $\sum_{k=1}^{\infty} Z_k$ 收敛，所以根据定理 2.15，我们就可以得到

$$n^{-1} \sum_{k=1}^n (Y_k - EY_k) \rightarrow 0 \quad \text{and hence} \quad \frac{T_n}{n} - n^{-1} \sum_{k=1}^n EY_k \rightarrow 0 \text{ a.s.}$$

分割线

考虑一系列独立同分布的随机变量 $\{X_n\}$ ，其中每个随机变量 X_n 服从均匀分布 $U(-n, n)$ 。这个序列的特征函数是

$$\phi_n(t) = \frac{\sin(nt)}{nt},$$

当 $n \rightarrow \infty$ 时， $\phi_n(t)$ 对于所有 $t \neq 0$ 都收敛到 0。但是，当 $t = 0$ 时， $\phi_n(0) = 1$ 对于所有的 n 都成立。

所以在这个例子中，我们可以说 $\phi_n(t)$ 在 $t = 0$ 处不收敛，而在 $t \neq 0$ 处收敛到 0。这种情况说明了特征函数的逐点收敛并不一定意味着对应的概率分布序列会弱收敛。

分割线